DOI:
10.4230/OASIcs.ICCSW.2012.149

Document Version
Publisher's PDF, also known as Version of record

# Introduction to Team Disruption Mechanisms

**Andrada Voinitchi[1], Elizabeth Black[1], and Michael Luck[1]**

1 **Department of Informatics, King's College London, UK**
andrada.voinitchi@kcl.ac.uk

## Abstract

This paper discusses how teams can be disrupted. More specifically, it discusses the steps that need to be taken in order to fully understand team disruption and design efficient mechanisms to disrupt teams. In order to answer the high-level question of how to disrupt teams, a few other questions need to be tackled first: what is a disrupted team? What are the crucial elements that make a collection of agents function as a team? Can norms, incentives or other mechanisms be used to disrupt these elements? How would we evaluate their efficiency? We first present the ideas of team and team disruption and motivate the need for these concepts to be properly defined. Secondly, we introduce an idea of team-disruption mechanism that we will further investigate. Lastly, we provide a long-term perspective and identify contributions that our research will make in the multi-agents field.

**1998 ACM Subject Classification** I.2. Computing Methodologies Artificial Intelligence

**Keywords and phrases** Team disruption, multi-agent systems, organisations, teams, goals

**Digital Object Identifier** 10.4230/OASIcs.xxx.yyy.p

## 1 Introduction

In order to better expose and motivate the question of how teams can be disrupted we consider a real-life scenario: a team of five terrorists are planning to place a bomb in a tube station. The bomb is heavy and needs to be smuggled in, part by part, in order for the station staff not to get suspicious. This can be done over time while keeping the parts hidden in the tube station or it can be done on the same day. Furthermore, the terrorists need to coordinate in order to assemble the bomb on the premises and leave the station before the bomb is detonated. What the terrorists do not know is that another team of undercover agents from the secret services have found out about their plan. We can say that the two teams compete, in the sense that only one of the teams can be successful in achieving its goal at a certain point in time:

- the terrorists want to blow up the station and bring about a state of the world where the specific station is destroyed;
- the secret services want to preserve the station and keep a state of the world where the specific station is intact.

There are two questions (of particular interest to us) arising from this example: how can a member of the secret services team infiltrate the terrorist team and disrupt their activity and, based on information provided by the secret services, how can the government introduce regulations in order to make it more difficult for terrorists to put bombs in tube stations?

This can also be applied in computational systems where teams of computational agents compete. It provides an insight into how team activity can be disrupted by a computational agent belonging to a competing team or by a legislator (the legislator is an agent that has the capability to introduce norms, which are rules that govern the behaviour of a computational


Conference/workshop/symposium title on which this volume is based on.
Editors: Billy Editor, Bill Editors; pp. 1–7
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

system). These two perspectives constitute a starting point for the process of designing and analysing team disruption mechanisms.

Team disruption is useful in various ways: when two teams compete, one of the teams may be at an advantage if they have mechanisms that can disrupt the other team's activity. Furthermore, team disruption is also useful when thinking about teams of humans: looking at our earlier example we can tell that a team disruption mechanism would be a good tool for the secret services to prevent a terrorist attack. Lastly, when it comes to team design, we need to ensure that we design our teams to be as efficient as possible. Knowing what can disrupt a team's activity we also know what to account for in the design, in order to make our teams more robust.

The contributions of this paper are two-fold: first, we provide a definition of team disruption; second, we suggest possible team disruption mechanisms that are to be investigated in future research.

The team disruption question discussed here is split into a set of preliminary questions that are discussed in detail throughout the paper. Section 2 focuses on a discussion of work that is relevant to this research such as a definition of teamwork and existing teamwork theories. Section 3 provides a breakdown of the team disruption question as well as a few ideas towards a methodology to be used in order to achieve team disruption. Section 4 comments on future work and its implications.

## 2 Related Work

While there has been much research conducted on teams and teamwork, as we discuss below, there is nothing specifically addressing the issue of team disruption.

Teamwork is a cooperative effort by the members of a team to achieve a common goal [9, 10]. It involves cooperative behaviour, coordination and at least one common goal among agents in the team. A team is formed by agents that agree to work together towards achieving a goal. However, between joining a team and achieving a goal there is a gap that needs to be resolved: how can agents work together in order to achieve the goal? Teamwork theories such as the Joint Intentions Theory [7], Joint Responsibility Theory [5, 6] and the Shared Plans Theory [3, 2] provide some answers to this question, describing mechanisms for achieving teamwork among agents. They are very relevant to our work because we need to consider how teamwork happens in order to be able to think about disrupting it.

## 3 Team disruption

This section is concerned with issues of what makes a team function and how it can be disrupted. In order to provide a better understanding of the technical concepts involved, preliminary definitions are provided.

### 3.1 Preliminary definitions

*Agent*: an agent is an autonomous entity that is characterised by its goal-directed behaviour, reactiveness to changes in its environment and its social ability (it is able to communicate with other agents). Two or more agents that interact within an environment form a *multi-agent system* (MAS). Furthermore, within a MAS, agents can work together towards achieving a common goal. This is referred to as a *team*: a team is a set of agents that share a common goal and cooperate in order to achieve it. This definition is further discussed in Section 3.2. In a MAS, agents may behave according to certain rules. These rules are called *norms*. In real life, norms are rules that govern a society and are used in order to regulate the behaviour of its individuals [8]. Norms in a MAS are the equivalent of written laws in a human society: agents need to comply with them in order to be rewarded or in order to prevent being punished. There are three kinds of norms: obligations, permissions and prohibitions [8] and only specific agents are allowed to introduce norms in a team. A MAS that supports the use of norms is also called a *normative MAS*.

## 3.2 What makes a team function?

A team is a set of agents that share a common goal. In a multi-agent setting, teamwork involves cooperative behaviour, coordination and communication [4]. Furthermore, teams are dynamic as agents can join or leave at any point in time.

Returning to the example mentioned in the introduction, the team of terrorists is a team because all of its members have one goal in common: destroy the tube station by detonating a bomb. Furthermore each of the terrorists has a role in the team: one terrorist is in charge of managing the others, some terrorists are in charge of carrying the bomb parts and so on.

It can be inferred from the previous definition of teamwork that, in order to function, a team has to support the following aspects: cooperation (the agents in a team need to act towards achieving the team goal instead of prioritizing their individual goals), coordination (to avoid redundancy: for example, to ensure that no two members attempt the same task simultaneously), communication (agents need to be able to communicate, as they need to coordinate and avoid duplicating work) and a common goal (if there is no common goal among all of the agents in a team then there is no reason for teamwork).

The aspects mentioned above are a few examples of what makes a team function. This remains a crucial question that has to be further researched in order to be able to identify what aspects can be sabotaged in order to disrupt a team.

## 3.3 What is a disrupted team?

The notion of a *disrupted team* is crucial for achieving an understanding of how teams can be disrupted. Team disruption, as we envision it, is based on the idea of rendering the team goal unachievable or causing a major delay in the achievement of the team goal (under the assumption that the team goal is initially achievable in a timely fashion). For example, in the context of our scenario, if one of the terrorists is observed carrying something suspicious by an undercover agent from the secret services team, then the terrorist has to hide, delaying the bombing, hence the team experiences disruption in the form of a delayed achievement of the team goal. If caught, the carrier terrorist is no longer able to deliver its part of the bomb, hence the goal is compromised and no longer achievable.

An understanding of what team disruption involves also relies on the way one thinks about reducing team efficiency. Is it enough to think of a disrupted team based on a delay in achieving the team goal or rendering the goal unachievable? How does this relate to lessening the number of members in the team? Does it imply disabling key members? It can also be assumed that finding ways of hindering any of the four aspects of a functioning team (cooperation, coordination, communication and common goal) can disrupt its activity.

A relevant aspect of team disruption is the agent that disrupts the team: a team can be disrupted by a malicious agent (possibly coming from another team) that has infiltrated within the team or through new regulations imposed by a legislator in the community that the team operates in.

When considering team disruption, we can consider not just whether it succeeds or fails but also the degree of success or failure, providing an idea of the extent to which a team can be disrupted. The concept of *magnitude of disruption* must be introduced when thinking about a disrupted team. This idea is at a very early stage and will be used in the future in order to evaluate team disruption mechanisms, thus it will not be discussed further here.

## 3.4 How to disrupt a team?

There are a few mechanisms that have the potential to disrupt teams (e.g. introducing norms, providing agents with incentives [1] to leave their team, reducing resources available to a team and so on). Here we will only address introducing norms in more detail, because of space restrictions.

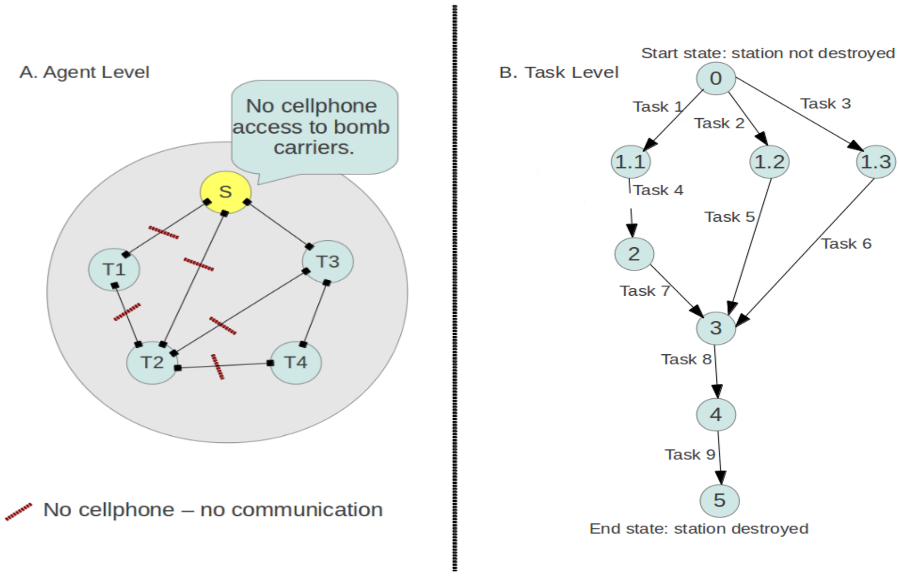


**Figure 1** *Disruption scenarios using norms: two perspectives for disrupting teams are illustrated both at agent level and at task level. In Figure 1.A. the nodes represent agents that are part of a terrorist team, while the links represent communication between agents. Agents T1-4 are members of the terrorist team while agent S is a member of the secret services that has infiltrated in the team. Figure 1.B. shows how the terrorist team goal is represented at task level. The nodes in the graph represent states of the world that the team operates in and the edges represent tasks that are accomplished in order to transition from a state to the other.*

When investigating team disruption, we can look at two levels of the team: the task level and the agent level (Figure 1). The agent level models the agent connections using a graph. The connections represent the ability of agents to communicate. The task level provides more detail on the sequence of steps towards achieving the team goal (task).

For example in Figure 1.A, agent S is an undercover agent infiltrating the terrorist team. In order to infiltrate the terrorist team, agent S declares to agent T3 (we assume agent T3 is in charge of managing the team) that it shares the goal of bombing a tube station. Once it joins the team, agent S gains the authority of setting norms for the team. It can then introduce a norm prohibiting agents that carry bomb parts (namely T1 and T2) from carrying cellphones, hence breaking their communication. This will make it more difficult for them to meet up and assemble the bomb. With no communication, agents T1 and T2 may not be able to find a common meeting point, leading to the goal being compromised. In the case where they meet accidentally, this may still delay the achievement of the goal as they may not meet in the place where they are supposed to assemble the bomb. This example illustrates team disruption at the agent level. From here we can infer that a team may be disrupted at the agent level by the introduction of norms that affect communication between agents that fulfill certain roles.

In Figure 1B, the same team of terrorists wishes to bomb a tube station. The topology of the team is as illustrated in Figure 1A. However, in this scenario, we do not focus on hindering communication for disrupting the team. This scenario relies on the fact that the team has a plan on how the bombing will be carried out. As shown in Figure 1B, the plan is composed of prioritized states that need to be reached in a specific order such that a connection between the start and the goal states is provided, as represented in a goal graph. In some cases, there may be more than one way of reaching the goal. The transition between states is realized through agent actions (tasks). Here, the states are represented as follows (we assume that the bomb is composed of three parts).

- 1.1: First bomb part is in the station.
- 1.2: Third bomb part is in the station.
- 1.3: Second bomb part is in the station.
- 2: Second bomb part is in the station.
- 3: All bomb parts are in the station.
- 4: Bomb is assembled (and in the station).

Tasks represent sets of actions performed by agents, as specified.

- Task 1: agent T1 carries first bomb part in the station.
- Task 2: agent T1 carries third bomb part in the station.
- Task 3: agent T1 carries second bomb part in the station.
- Task 4: agent T2 carries second bomb part in the station.
- Task 5: agent T2 and T3 carry the first and second bomb parts in the station.
- Task 6: agent T2 and T3 carry the first and third bomb parts in the station.
- Task 7: agent T3 carries the third bomb part into the station.
- Task 8: agent T2 assembles the bomb.
- Task 9: agent T4 detonates the bomb.

As shown in the example in Figure 1B, more members of the team can either carry the bomb parts simultaneously, meet in the station and assemble the bomb (nodes 1.2 and 1.3 of the goal graph), or one member can carry all of the bomb parts, part by part, store them in the station and assemble it when all of the parts have arrived (node 1.1. of the goal graph). This example is different from the previous one (Figure 1A) in that it does not imply direct sabotage by an agent infiltrating. Rather, agents from a competing team can use their observations to convince legislators to introduce norms: if it is known that bomb parts can be stored unattended in tube stations, the legislators (i.e. government) can introduce a law (norm) stating that no bags can be left unattended in the tube station. Unattended bags are seized and destroyed. Because State 2 from the goal graph is no longer achievable, the team may be delayed in achieving the goal (bombing) because their options were reduced. Furthermore, if legislators introduce a norm restricting bag size for travelers, larger parts of the bomb may not be introduced into the station at any time, thus the bombing is fully compromised.

When considering disruption at task level, a few issues need to be considered for an efficient approach, as follows.

- First, in order to render the goal unachievable, all paths to the goal need to be compromised (in the goal graph).
- Second, some tasks may be crucial to accomplishing the team goal: if these tasks are not accomplished, the goal is rendered unachievable (such as Task 8 in the example in Figure 1B). These are considered *critical tasks* and the agents that can perform them are deemed critical agents.

## 4 Conclusion and future work

Based on the results that will be obtained from the planned simulation of all of the mechanisms presented, a further step is to be taken, comparing the outcomes of the scenarios based on criteria such as the time it takes a team to achieve a goal and if the team can achieve its pre-disruption goals. A further investigation will establish whether each of these mechanisms can work individually or whether applied together they will have a greater impact.

If it is demonstrated that any of the methods described here can be used in order to disrupt team activity further questions need to be answered. For example, if introducing norms can be responsible for team disruption, a number of issues need to be further investigated: how would such norms be generated? Who would introduce and enforce such norms? What proportion of the agents in a team would need to comply with such norms in order for the team to be disrupted? If all of the members of a team need to comply with the norms, who would supervise the norm enforcers in order to assure compliance? As a further step, it is vital to compare all the successful approaches and provide a detailed analysis on which of these approaches work better and under which circumstances.

## References


1. Y. Chen, J. Kung, D.C. Parkes, A.D. Procaccia, and H. Zhang. Incentive design for adaptive agents. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '11, pages 627–634, Richland, SC, 2011. International Foundation for Autonomous Agents and Multiagent Systems.
2. B.J. Grosz, L. Hunsberger, and S. Kraus. Planning and acting together. *AI Magazine*, (20):23–34, 1999.
3. B.J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, October 1996.
4. B. Horling and V. Lesser. A survey of multi-agent organizational paradigms. *Knowledge Engineering Review*, 19(4):281–316, December 2004.
5. N.R. Jennings. On being responsible. *SIGOIS*, 13(3):8–, December 1992.
6. N.R. Jennings. Commitments and conventions: The foundation of coordination in multi-agent systems. *Knowledge Engineering Review*, 3(8):223–250, 1993.
7. H.J. Levesque, P.R. Cohen, and J.H.T. Nunes. On acting together. In *The 8th National conference on Artificial intelligence - Volume 1*, AAAI'90, pages 94–99. AAAI Press, 1990.
8. S. Modgil, N. Faci, F. Meneguzzi, N. Oren, S. Miles, and M. Luck. A framework for monitoring agent-based normative systems. In *The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '09, pages 153–160, Richland, SC, 2009. International Foundation for Autonomous Agents and Multiagent Systems.
9. D.V. Pynadath and M. Tambe. Multiagent teamwork: analyzing the optimality and complexity of key theories and models. In *The 1st international joint conference on Autonomous agents and multiagent systems: part 2*, AAMAS '02, pages 873–880, New York, NY, USA, 2002. ACM.
10. M. Tambe. Towards flexible teamwork. *Journal of Artificial Intellience Research*, 7(1):83–124, September 1997.